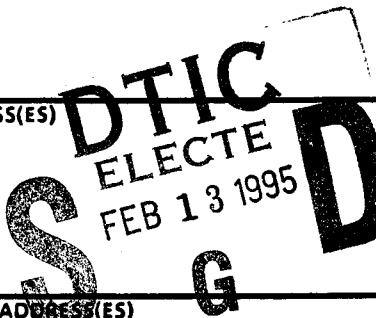


REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 9/20/94	3. REPORT TYPE AND DATES COVERED FINAL 1 Aug 92 - 31 Jul 94		
4. TITLE AND SUBTITLE UNCONVENTIONAL IMAGE RECOVERY TECHNIQUES		5. FUNDING NUMBERS DAAL03-92-C-0045		
6. AUTHOR(S) RICHARD BARAKAT				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) RGB ASSOCIATES, INC. P.O. BOX 8 WAYLAND, MA 01778				
8. PERFORMING ORGANIZATION REPORT NUMBER		10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARO 30606.1-MA-SDI		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709-2211		11. SUPPLEMENTARY NOTES The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.		
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.		12b. DISTRIBUTION CODE		
13. ABSTRACT (Maximum 200 words) The main theme of this contract is the development of unconventional imaging techniques. The imaging methods advocated result in complicated nonlinear relations between the object and its image, inversion techniques are required to translate the measured data into a meaningful estimate of the object. Although these methods require a large amount of computation, they are relatively cheap compared to the costs of large aperture optical systems. The technical approach is two-fold: first development of the theory of the unconventional imaging scenario in question; second adaption of the latest techniques from numerical analysis to carry out the inversion from measured data to reconstructed object. One problem was completed and written up for publication: Inversion of the modulus and phase of a coherently illuminated object from its measured diffraction image. This constitutes the main body of the report along with representative numerical calculations. The inversion techniques developed here should be of great use in various biomedical imaging techniques and could result in replacing phase contract microscopy.				
14. SUBJECT TERMS Diffraction theory, inverse problems		15. NUMBER OF PAGES 33		16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

19950203 314

DTIC QUALITY ASSURANCE

UNCONVENTIONAL IMAGE RECOVERY TECHNIQUES

FINAL REPORT

RICHARD BARAKAT
02 OCTOBER 1994

U.S. ARMY RESEARCH OFFICE

DAAL03-92-C-0045

RGB ASSOCIATES

APPROVED FOR PUBLIC RELEASE

DISTRIBUTION UNLIMITED

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input checked="checked" type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

THE VIEWS, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

The requirements on SDIO imaging problems, although somewhat severe, can generally be accomplished by using very large optics. Unfortunately the expense associated with constructing, testing and deployment is generally prohibitive (as I can attest having served on the committee investigating the Hubble Space Telescope failure) and other, less expensive approaches, are desirable.

The main theme of this contract is the development of unconventional imaging techniques. The various unconventional imaging methods advocated result in complicated nonlinear relations between the object and its image, inversion techniques are required to translate the measured data into a meaningful object. Although these methods require a large amount of computation, they are relatively cheap compared to the costs of large telescopes which rely upon physical size for increased resolution.

The technical approach is two-fold: first, development of the theory of the unconventional imaging scenario in question; second, adapt the latest techniques from numerical analysis to carry out the inversion from measured data to reconstructed object. Any inversion technique adopted must be robust with respect to measurement noise.

The contract was originally for two years, but only one year was actually funded. As a consequence only one of the three research projects was completely finished. This work has been written up for publication under the title "Inversion of the Modulus/Phase of a Coherently Illuminated Object from its Measured Diffraction Image". A copy of the manuscript is included in this section. Aside from its general significance, these techniques developed in this paper are of direct use in optical microscopy and can be further developed so as to supplement the well known technique of phase contrast microscopy. The inversion algorithm developed here can also be used for lens testing and could, with some further work, replace the cumbersome knife edge

test for wavefront aberrations.

Forward

Table of Contents

Main Body of Report: Inversion of the modulus and phase
of a coherently illuminated object from its measured
diffraction image.

References

Figures with Figure Legends

Report of Inventions

FIGURE LEGENDS

Fig. 1 Distribution of illuminence in the diffraction image of a coherently illuminated opaque edge: $o(v) = 0, v < 0$ and $o(v) = 1$ for $v > 0$ viewed through an annular aperture of obscuration radius $\epsilon = 0.05$. The solid line is for the in-focus situation, the dotted line is for a half-wave of defocus. Taken from Reference 5.

Fig. 2 Sample realization of reconstruction of modulus and phase of coherently illuminated object (dashed lines) in the presence of 2% measurement "noise".

Fig. 3 Sample realization of reconstruction of modulus of coherently illuminated object (dashed line) in the presence of 4% measurement "noise".

Fig. 4 Sample realization of reconstruction of modulus of coherently illuminated object (dashed line) in the presence of 4% measurement "noise".

**Inversion of the Modulus/Phase of a Coherently Illuminated
Object from its Measured Diffraction Image**

ABSTRACT

An algorithm for recovering the modulus and phase of a coherently illuminated object from its measured diffraction image is presented. The algorithm is based upon the fact that both the Jacobian and Hessian matrices can be evaluated exactly so that both slope and curvature information is available. The inversion problem is cast as a nonlinear unconstrained optimization problem, and trust region techniques are employed for its solution. Representative numericals are presented.

1. INTRODUCTION

The problem of measuring the diffraction image of an incoherently illuminated object and working back to determine (or estimate) the object itself has been of great interest in many scientific and technical areas for the past twenty years starting with the work of Barakat and Blackman [1] who employed Tichonov regularization methods. The subject of object restoration for incoherently illuminated objects has grown to such an extent that even listing the books and major review articles is a tedious undertaking; however, we should refer to the standard work of Andrews and Hunt [2] which has served to educate so many people interested in the topic.

At the other extreme, we have the analogous problem for a coherently illuminated object, a much more difficult problem because the relation between object and diffraction image is nonlinear, whereas the corresponding relation for incoherently illuminated objects is linear. In many respects the coherently illuminated situation is much older than the incoherently illuminated situation in that light microscopists have always encountered such problems. Their "solutions" have generally been of the old-fashioned expert systems type; they have encountered over the years a variety of biological objects and developed an empirical expertise in sorting out situations. In a sense their primary artifact is the diffraction image, not the actual object, as witness the various phase contrast methods [3,4]. Unlike the incoherent situation, there is no guarantee that the topology of the object bears any resemblance to its coherent diffraction image. As an example see Figure 1 which shows the diffraction image of an edge viewed through an optical system with an annular aperture [5].

In the present paper a solution for the modulus and phase of a coherently illuminated object is obtained via a nonlinear regularized minimization. We have brought to bear powerful tools recently developed in numerical analysis (particularly trust region considerations) toward the efficient solution of the inversion problem. Even with a CRAY, the

illustrative calculations at the end of the paper required running times of up to several minutes.

The present scheme takes advantage of the fact that we can evaluate both Jacobian and Hessian matrices *exactly*, thereby allowing us to make use of slope and curvature information explicitly. There is no need to make the usual small residual approximation of the Hessian with its convergence limitation in the presence of measurement noise. A second benefit of knowing the Jacobian and Hessian explicitly is that we can make very efficient use of the trust region tests in determining the path to local minima.

There are a number of algorithms for the inversion of modulus/phase of coherently illuminated objects already published; see Stark [6] for an extremely useful summary. Practically all these algorithms will yield a reasonably accurate estimate of the support of the object, and the present algorithm is no exception. However, many coherently illuminated objects contain changes in modular/phase over the surface so that an algorithm (such as the one discussed here) which also can deliver information on the distribution of modulus/phase over the object is extremely useful for applications (e.g., biological light microscopy).

2. DIFFRACTION MODEL

The diffraction image of a coherently illuminated object, $I(x, y)$ in the image receiving plane with coordinates x, y is measured over a square lattice of points x_m, y_n :

$$\begin{aligned} x_m &= \beta_m \\ y_n &= \beta_n \end{aligned} \quad m, n = 0, \pm 1, \pm 2, \dots, \pm M \quad (2.1)$$

where β is a numerical constant. It is assumed that M is large enough for

$$I(x_{|M|}, y_{|M|}) \approx 0 \quad (2.2)$$

In what follows, it will be convenient to write $(x_m, y_m) \equiv h_{mn}$ as a column vector \mathbf{I} of length $\tilde{m} \equiv (2M + 1)^2$ using standard Fortran lexicographic ordering.

We assume that the measured diffraction image can be modelled via scalar optical diffraction theory so that the model diffraction image, $\mathcal{I}(x, y)$, is given by the convolution

$$\mathcal{I}(x, y) = \left| \int \int_{\text{object}} a(x - x', y - y') o(x', y') dx' dy' \right|^2 \quad (2.3)$$

assuming the isoplanatic condition to hold.

The coherently illuminated object is characterized by a complex-valued function $o(x, y)$ with modulus $|o(x, y)|$ and phase $\arctan[o_i(x, y)/o_n(x, y)]$. The function $a(x, y)$ is the coherent point-spread function of the optical system performing the imaging; to within multiplicative factors it is

$$a(x, y) = \int \int_{\text{exit pupil}} A(\zeta, \eta) e^{i\frac{\lambda}{2}(x\zeta + y\eta)} d\zeta d\eta \quad (2.4)$$

with k as the mean wavenumber of the coherent light and f the focal length. The pupil function $A(\zeta, \eta)$ is given by

$$A(\zeta, \eta) = B(\zeta, \eta)e^{ikW(\zeta, \eta)} \quad (2.5)$$

with $W(\zeta, \eta)$ the wavefront aberration function over the exit pupil and $B(\zeta, \eta)$ the amplitude distribution over the exit pupil; both are assumed known in the present scenario.

For the very important case of a circular aperture exit pupil of radius α for which $B(\zeta, \eta) = 1$ and $W(\zeta, \eta) = 0$, we have

$$a(x, y) = \frac{2J_1 \left[\frac{k\alpha}{f}(x^2 + y^2)^{1/2} \right]}{\left[\frac{k\alpha}{f}(x^2 + y^2)^{1/2} \right]} \quad (2.6)$$

Since in most applications, the exit pupil is circular; under these circumstances it is convenient to employ normalized coordinates:

$$\begin{aligned} u &\equiv \frac{k\alpha}{f}x, & v &\equiv \frac{k\alpha}{f}y \\ p &= \frac{\zeta}{\alpha}, & q &= \frac{\eta}{\alpha} \end{aligned} \quad (2.7)$$

Consequently Eqs. (2.3), (2.4), and (2.6) become

$$\mathcal{I}(u, v) = \left| \int \int_{\text{object}} a(u - u', v - v') o(u', v') du' dv' \right|^2 \quad (2.3a)$$

$$a(u, v) = \iint_{0 \leq p^2 + q^2 \leq 1} A(p, q) e^{i(u p + v q)} dp dq \quad (2.4a)$$

$$a(u, v) = \frac{2J_1((u^2 + v^2)^{1/2})}{(u^2 + v^2)^{1/2}} \quad (2.6a)$$

Leaving aside the mathematical details for the moment, let us consider the task before us. We are given *measured* values of the diffracted intensity and are required to determine the modulus and phase of the object assuming that the measured diffracted intensity can be modeled as the convolution, Eq. (2.3), and that we know the parameters of the optical system characterizing the coherent point spread function, Eq. (2.4). Said object is contained in the integrand of a double integral which is itself squared. When viewed from this perspective, we should not be unduly optimistic about achieving an accurate solution because of the inherent nonlinearity and strong smoothing action of the double integral. Irrespective of the actual inversion method, the strong smoothing action of the integration means that much high frequency data is lost and cannot be retrieved, even in principle. Only a low frequency version of the object can be obtained. Perhaps the best way to consider the problem is to interpret it thusly: we are given the "answer" (\equiv effect) in the form of a noisy diffraction image of the object and are attempting to determine the "question" (\equiv cause), the coherently illuminated object. The mathematical relation between answer and question is highly nonlinear by virtue of Eq. (2.2), bearing in mind that $a(u, v)$ is additionally an oscillating, complex-valued function.

In a sense this problem can be considered as a two-dimensional phase retrieval problem, see [7-13] for various details. We will not discuss the general issues of ill-posed inverse problems and refer the reader to [14,15] for *some* of the theoretical aspects.

Our diffraction model image $\mathcal{I}(u, v)$ is to be determined by the values of the image at the lattice points (u'_k, v'_l) inside a square in the (u', v') plane that is large enough to contain the image. The number of object values, $o(v'_k, v'_l) \equiv o_{kl}$ is taken to be N . It is important to remember that both modulus and phase must be determined at each of these lattice points. Let \tilde{n} be the number of unknown modulus and phase values (obviously \tilde{n} must be even); the unknown o_{kl} are to be written as a column vector \mathbf{o} .

In this notation Eq. (2.3a) becomes $\mathcal{I}(\mathbf{o})$ where \mathcal{I} is a vector of length \tilde{m} . The inversion problem relates the unknown (complex-valued) object values \mathbf{o} of the assumed diffraction model to the measured diffraction image \mathbf{I} via the nonlinear relation

$$\mathbf{I} = \mathcal{I}(\mathbf{o}) \quad (2.7)$$

This equation is to be interpreted as a system of \tilde{m} nonlinear equations in \tilde{n} unknowns. Enough data must be given to allow some smoothing of the measured diffraction image data as regards the diffraction model; consequently we let $\tilde{m} > \tilde{n}$ so that the nonlinear system is overdetermined. The problem now reduces to the solution of Eq. (2.7) in some normed sense.

It is of some interest to contrast the differences between the incoherently illuminated and coherently illuminated object situations. The relation between object and image for incoherently illuminated objects, again assuming that the isoplanatic conditions holds, is

$$I_i(x, y) = \int \int_{\text{object}} t(x - x', y - y') o_i(x', y') dx' dy' \quad (2.8)$$

where

$o_i(x, y)$ = intensity distribution over the
incoherently radiating object

$t(x, y) \propto |a(x, y)|^2$ = incoherent point-spread function

$I_i(x, y)$ = intensity distribution over the diffraction
image of $o_i(x, y)$

Two points to note: (1) all functions in Eq. (2.8) are real and nonnegative, (2) the relation between answer, $I_i(x, y)$, and question, $o_i(x, y)$, is linear. On the other hand, for the coherently illuminated object only the diffracted intensity is real and nonnegative. Furthermore the relation between $I(x, y)$ and $o(x, y)$ is nonlinear. Consequently the incoherently illuminated object scenario is much easier to invert than the coherently illuminated object scenario.

3. PRELIMINARIES

In order to proceed, we next discretize the double integral on the right-hand side of Eq. (2.3a)

$$\mathcal{I}(u_m, v_m) \equiv \mathcal{I}_{mn} = \left| C \sum_{k=1}^N \sum_{\ell=1}^N \alpha_k \alpha_\ell a(u_m - u_k, v_n - v_\ell) o(v_k, v_\ell) \right|^2 \quad (3.1)$$

Here u_k, v_ℓ are the quadrature points and α_k, α_ℓ are the corresponding weight factors.

In the numerical algorithm we employ for the inversion (see next section), we require the first and second derivatives of \mathcal{I}_{mn} with respect to $o_{k\ell}$ in order to form Jacobian and Hessian matrices. We need not write out the explicit formulas because we employed a symbol manipulation program to evaluate them directly in the computer from which numerical values are obtained internally.

4. OUTLINE OF SOLUTION APPROACH

We now outline the *general* features of our numerical approach to the nonlinear phase retrieval problem via regularized unconstrained minimization. Our version of the unconstrained minimization problem is that of finding the least value of an objective function, $\epsilon(o)$. The term unconstrained indicates that the variables o are not limited in any way. In our application, we wish to determine a global minimum of ϵ , i.e., a point o^* satisfying

$$\epsilon(o) \geq \epsilon(o^*), \quad \forall o \quad (4.1)$$

Unfortunately we must be content with a local minimum:

$$\epsilon(o) \geq \epsilon(o^*) \quad \forall o, \text{ in a neighborhood of } o^* \quad (4.2)$$

Solution of the global minimum problem is far harder than the solution of the local minimum problem which itself is extremely complicated [16,17].

The derivation of nearly all methods for unconstrained minimization is founded on the assumption that the objective function $\epsilon(o)$ can be approximated by a quadratic function in the neighborhood of a minimum point. Thus methods are sought which efficiently minimize quadratic functions in the hope that they will also be effective on more general functions, at least in the neighborhood of a minimum. When first and second derivatives of $\epsilon(o)$ are known, such as in our case, then we can make use of both gradient and curvature information to effect a solution. Furthermore, both gradient and curvature, as governed by G and H , see Eqs. (4.4) and (4.5), can be calculated *exactly* in the context of the discretized version of our problem. Thus, we can avoid many of the difficulties associated with situations where G and H are known only approximately. The Taylor series expansion of the objective function can be used to approximate its minimum value from points o near to the minimum o^* by moving in a direction Δo

$$\epsilon(o + \Delta o) \approx \epsilon(o) + G^T(o)(\Delta o) + \frac{1}{2}(\Delta o)^T H(o)(\Delta o) \quad (4.3)$$

G and H are the gradient vector and Hessian matrix of the objective function, respectively

$$G^T \equiv \left[\frac{\partial \epsilon}{\partial o_1}, \frac{\partial \epsilon}{\partial o_2}, \dots, \frac{\partial \epsilon}{\partial o_n} \right] \quad (4.4)$$

$$H \equiv \begin{bmatrix} \frac{\partial^2 \epsilon}{\partial o_1^2} & \dots & \frac{\partial^2 \epsilon}{\partial o_1 \partial o_n} \\ \vdots & & \vdots \\ \frac{\partial^2 \epsilon}{\partial o_n \partial o_1} & \dots & \frac{\partial^2 \epsilon}{\partial o_n^2} \end{bmatrix} \quad (4.5)$$

Note that H is symmetric. Both G and H are exact within the context of the discretization. The strategy is to determine the vector Δo of the movements required to approximate the minimum from the current point o .

Before proceeding further, it is important to state the conditions for a given point o to be an unconstrained strong minimum o^* (i.e., a point o^* for which the objective function increases locally in all directions. The first order necessary condition is [16,17]

$$G(o^*) = 0 \quad (4.6)$$

However, this condition is not sufficient as other types of minima (stationary points) also satisfy this condition. The second order condition follows from Eq. (4.2) and is [11,12]

$$(\Delta o)^T H(o^*) (\Delta o) > 0 \quad (4.7)$$

which is sufficient to ensure that $\epsilon(o^* + \Delta o) > \epsilon(o^*)$. If $\epsilon(o) \neq 0$, this implies that $H(o^*)$ is a positive definite matrix. Therefore the second-order sufficient condition for a strong minimum is that $H(o^*)$ should be positive definite.

Returning to Eq. (4.3), hereafter termed the local model of the object function, we initiate a search for o^* by moving Δo . This involves an iterative procedure in that we

start from some point \mathbf{o} and choose in some fashion a direction $\Delta\mathbf{o}$ in which the minimum is assumed to lie. This is repeated until the minimum is achieved (if possible). Because we are employing a quadratic version of $\epsilon(\mathbf{o})$ then the local model of $\epsilon(\mathbf{o})$ always allows us to find a solution. However, the local model of $\epsilon(\mathbf{o})$ is certainly not a useful approximation to $\epsilon(\mathbf{o})$ itself except near a minimum. It is an act of faith that the local model and the actual model are "close" in the vicinity of the minimum. Simple calculations [16,17] show that at a minimum

$$\Delta\mathbf{o} = -[\mathbf{H}(\mathbf{o})]^{-1}\mathbf{G}(\mathbf{o}) \quad . \quad (4.8)$$

This equation represents the appropriate direction $\Delta\mathbf{o}$ to take to the minimum \mathbf{o}^* , based upon information at \mathbf{o} . This equation is fundamental to all second-order minimization algorithms (i.e., algorithms employing both \mathbf{G} and \mathbf{H}). However, on the actual objective function surface, such as we are faced with, the local model of $\epsilon(\mathbf{o})$ is only accurate in the *immediate* vicinity of the minimum. This means that \mathbf{H} may not be positive definite, as required by Eq. (4.7), since it is evaluated at points other than the minimum. This situation is most likely to occur at some distance from the minimum (i.e., for initial iterations) since at a point close to the minimum all sufficiently differentiable functions tend to behave as a quadratic function as their third and higher order derivatives in the Taylor series become negligible. We will return to the necessity of keeping \mathbf{H} positive definite very shortly.

The classic approach to limiting step size during the iteration is a line search [16,17]. In line search tactics we compute a descent direction and subject this direction (which is generally not toward the unconstrained minimum) to a minimization procedure of which details can be found in the above references. Should the descent direction satisfy these criteria, this iteration is then terminated. We then repeat the process, etc. The difficulty of practical implementation is two-fold. First, such calculations are prohibitively expensive and the minimization criteria are therefore only approximate to save computer time;

yet they must be made precise enough to ensure reasonably quick convergence. Second, fulfilling these contradictory goals is something of a black art and the programming effort is very substantial, often occupying up to two-thirds of the coding for the entire optimization. For small problems, such as the slit aperture, line search methods are very useful and were employed along with regularized singular value decomposition in [18].

Although it is possible to use line search methodology for the 2-D aperture, we have chosen to employ a relatively new method which offers many advantages over the line search methodology, the trust region tactic [16.7], but see also [18]. A particular advantage of the trust region approach is in the very strong global convergence properties which hold with no significant restrictions on the class of problems to which they apply making it especially useful for the nonlinear minimization of phase retrieval in two dimensions. In addition, the trust region philosophy requires less computation than does the line search philosophy in terms of gradient and Hessian, but more computations have to be performed on the local model of the objective function.

Suppose we are at point $\mathbf{o}^{(k)}$, after the k -th iteration. Now let us *assume* that there is a region $\Delta^{(k)}$, which we take to be in the shape of a sphere of radius $h^{(k)}$

$$\Delta^{(k)} = \{\mathbf{o} : \|\mathbf{o} - \mathbf{o}^{(k)}\| \leq h^{(k)}\} \quad (4.9)$$

in which the local model of $\epsilon(\mathbf{o})$, given by the Taylor series Eq. (4.2) agrees with the actual objective function in some sense. One's obvious choice is to let

$$\mathbf{o}^{(k+1)} = \mathbf{o}^{(k)} + \Delta^{(k)} \quad (4.10)$$

where the correction $\Delta^{(k)}$ minimizes the local model (Δ) for all values of $\mathbf{o}^{(k)} + \Delta$ in $\Delta^{(k)}$. An iteration step in \mathbf{o} is to be restricted by the region of validity of the Taylor series. Thus we compute the gradient and Hessian, G and H , appropriate to $\mathbf{o}^{(k)}$ and minimize the

local model of $\epsilon(o)$ in order to determine the radius $h^{(k)}$ of the trust region. In formal language, we seek the solution of the problem

$$\text{minimize (local model) subject to constraint } \|\Delta\|_2 \leq h^{(k)}. \quad (4.11)$$

Before we can solve this subproblem it is necessary to choose some reasonable criterion on the radius $h^{(k)}$ of the trust region. Obviously, the criterion should not present undue restrictions, so that $h^{(k)}$ should be as large as possible subject to some agreed-upon idea as to the degree of closeness of the local model of $\epsilon(o)$ and $\epsilon(o)$. One way is to define the quality coefficient

$$r_k \equiv \frac{\epsilon(o^{(k)} + \Delta^{(k)}) - \epsilon(o^{(k)})}{\epsilon_q(o^{(k)} + \Delta^{(k)}) - \epsilon_q(o^{(k)})} \quad (4.12)$$

where the denominator is the predicted reduction of the local model of ϵ , now call it ϵ_q , while the numerator represents the reduction in the actual objective function. The closer r_k is to unity, the better the agreement between ϵ and ϵ_q . As a stopping criterion stop if $r_k > .8$ and go to the next iteration. If $r_k < .8$ reduce $h^{(k)}$ and repeat the calculation. The literature contains other stopping criteria but the one quoted above seems adequate for our purposes. See the quoted references for more details.

Thus far we have outlined the general features of the inversion problem, estimating o via minimization of an objective function ϵ , as yet unspecified. We feel that there are two objective functions of possible interest. The first is the usual least squares (L_2 norm) objective function

$$\epsilon(o) = \sum_{i=1}^m (I_i - \mathcal{I}_i)^2 \quad (4.13)$$

In this version of the problem, we consider the estimation of o via an (unconstrained) nonlinear least squares minimization. A second objective function is for an L_1 norm

$$\epsilon(o) = \sum_{i=1}^m |I_i - \mathcal{I}_i| \quad (4.14)$$

Consequently we have allowed considerable flexibility in the ϵ minimization algorithm so as to accommodate both objective functions as well as others that may arise. For the purposes of the present paper, we confine the discussion to the L_2 norm aspects.

Upon defining the vector

$$\Phi(o) = \mathcal{I} - \mathbf{I} \quad (4.15)$$

which is of length \tilde{m} , we can rewrite Eq. (4.13) as

$$\epsilon(o) = \Phi^T(o)\Phi(o) \quad (4.16)$$

The elements of G and H can be expressed directly in terms of the elements Φ by introducing an ancillary matrix J given by

$$J \equiv \begin{vmatrix} \frac{\partial \phi_1}{\partial o_1} & \dots & \frac{\partial \phi_1}{\partial o_{\tilde{n}}} \\ \vdots & & \vdots \\ \frac{\partial \phi_{\tilde{m}}}{\partial o_1} & \dots & \frac{\partial \phi_{\tilde{m}}}{\partial o_{\tilde{n}}} \end{vmatrix} \quad (4.17)$$

This matrix is generally rectangular. Thus

$$\frac{\partial \epsilon}{\partial o_j} = 2 \sum_{i=1}^{\tilde{m}} \phi_i \frac{\partial \phi_i}{\partial o_j} \quad (4.18)$$

or

$$G = 2J^T \Phi \quad (4.19)$$

The corresponding elements of H are

$$\frac{\partial^2 \epsilon}{\partial o_k \partial o_l} = 2 \sum_{i=1}^{\tilde{m}} \frac{\partial \phi_i}{\partial o_k} \frac{\partial \phi_i}{\partial o_l} + 2 \sum_{i=1}^{\tilde{m}} \phi_i \frac{\partial^2 \phi_i}{\partial o_k \partial o_l} \quad (4.20)$$

with $k, \ell = 1, 2, \dots, \tilde{n}$. The first term on the right-hand side is $2J^T J$ and we will call the second term S ; thus

$$H = 2J^T J + S \quad (4.21)$$

There is no guarantee that H will always be positive definite through the calculations; in fact H will become almost singular as we approach the strong minimum. To avoid this state of affairs, we consider a regularized version of H . In the regularized version, we replace H by $H + qI$ where q , the regularizing parameter, is a small nonnegative real number. This procedure can remove H from near singularity and for sufficient large q restore the positive definite character of the Hessian. In linear problems, there is a well established method to estimate q [14]. In the nonlinear problem we employed q in the range $0.01 \leq q \leq 0.05$. There was not much difference in the final answers as long as $q > 0$; but setting $q = 0$ caused considerable numerical instability as expected.

5. A NUMERICAL EXAMPLE

It is not our intention to present a catalog of numerical results at this time, yet we wish to discuss the numerical workings of the algorithm in the presence of "measurement" noise. To this end we considered a known object $o(u, v)$ and from it calculated the diffraction image $I(u, v)$ from Eq. (2.3a) using the point-spread function $a(u, v)$ corresponding to an in-focus, aberration-free optical system as given by Eq. (2.6a). The *measured* diffraction image $I(u, v)$ was taken to be given by

$$I(u, v) = [1 + \delta\mu(u, v)]I(u, v) \quad (5.1)$$

where δ is a positive constant less than unity and $\mu(u_m, v_n)$ is a random variable uniformly distributed over $(-1, +1)$:

$$\begin{aligned} f(\mu) &= \frac{1}{2}, & |\mu| < 1 \\ &= 0, & |\mu| > 1 \end{aligned} \quad (5.2)$$

Values of δ used in the present calculations, such as $\delta = 0.04$ are described loosely as 4% a noise. Note that the noise we are introducing is intensity dependent noise; it is not intended to faithfully simulate actual detector noise but rather to mimic such noise for the purpose of testing the robustness of the inversion algorithm.

We have chosen the following lone object to illustrate the calculations:

$$\begin{aligned} o(u, v) &= 0, & -\infty < v < -15 \\ &= \frac{3}{8}[3 + \text{erf}(v)], & -15 < v < 15 \\ &= 0, & -15 < v < \infty \end{aligned} \quad (5.3)$$

where $\text{erf}(v)$ is the error function. There are two reasons for choosing this object. The first reason is that the modulus of the object exhibits a spatial variation. One of the

(as yet unstated) requirements on the algorithm is that it be able to recover reasonably accurate estimates of the object photometry in addition to estimating its size; as noted in the introduction almost any of the currently available algorithms can perform this; very few seem to be able to provide reasonably accurate estimates of photometry. The second reason for choosing this object is that it is very small in width being slightly less than Airy disk radius across (recall that an Airy disk radius ≈ 3.82). These two constraints present a severe test of the algorithm.

Calculations were run for the *completely unrealistic* case of the noise-free, measured diffraction image of the object given by Eq. (5.3). The inversion algorithm was able to return reasonable answers as compared to the true values. We will not quote any of these results and go to the "noisy" measurement situation.

In Figure 2 we show a sample realization of the reconstruction in the presence of 2% measurement noise. Note the presence of negative values of the modulus at the edges of the object; however the photometry of the reconstruction follows the true object very well except at the edges. Two sample realizations of the reconstruction of the modulus of the object, in the presence of 4% measurement noise are shown in Figs. 3 and 4. Again note the unphysical negative values of the modulus at the edges of the object. Considering the fact that the object is so small, the modulus is in reasonable agreement with the true object in spite of measurement noise.

The troublesome feature of the present inversion algorithm is obviously the occurrence of negative modulus values at the edges of the object. The edges are somewhat unrealistic because they are of infinite slope and the algorithm tends to overshoot in the manner of a Gibbs phenomenon. We could, of course, consider objects with finite slopes at the edges to minimize the overshoots and undershoots; instead we have decided to develop a variant of the inversion algorithm using constrained minimization to surmount this annoying problem. Details will be reported in the near future.

Although we have run several dozen inversions, we do not possess a sufficiently large base to estimate benchmark performance. However, many of our inversions, regardless of the initial guess, took 150-180 iterates to converge. In a few cases, convergence was achieved in half this number; nevertheless we also encountered some cases where convergence required 200 iterations.

Generally the algorithm performance decomposed into two stages: (a) the initial stage wherein the objective functions went through wild gyrations in magnitude as the trust region subalgorithm had to compute many new gradient and Hessian matrices while searching for an appropriate direction of descent; (b) the second stage, when the trust region subalgorithm has locked into a subset of "optimal" directions of descent, the objective function then undergoes a monotone decrease to zero as the number of iterations is increased. We again warn the reader that although convergence is attained, there is no guarantee that the global minimum has been achieved.

As a final remark, we note that as the analysis is heavily dependent upon linear algebra (matrices, vectors); it would be of some interest to consider doing calculations via parallel processing algorithms to speed up the computations even more.

6. SUMMARY

Given the rather discursive presentation, we feel it is useful to summarize the essentials of our approach to the problem. We consider the problem of determining the modulus and phase of the object at the lattice points as one of unconstrained minimization. A local (i.e., quadratic) model approach is used. We are in the fortunate position of employing both the Jacobian and Hessian, which, in the context of the discretized diffraction model, can be evaluated analytically! Thus we make use of both slope and curvature information, other methods only use slope information. The relatively new, and powerful trust region algorithm is then utilized in place of the usual line search algorithm. The trust region algorithm makes full use of the local model and takes into account local validity; moreover the algorithm exhibits global convergence properties.

REFERENCES

1. R. Barakat and E. Blackman, "Application of the Tichonov regularization algorithm to object restoration," *Opt. Comm.* 9, 252-256 (1973).
2. H. Andrews and R. Hunt, *Digital Image Restoration* (Prentice Hall, Englewood Cliffs, NJ, 1977).
3. A.H. Bennett, H. Jupnik, H. Osterberg, and O.W. Richards, *Phase Contrast Microscopy* (Wiley, New York, 1951).
4. M. Pluta, *Advanced Light Microscopy*, vol. 2 (Elsevier, Amsterdam, 1988), Chpt. 5.
5. R. Barakat, "Diffraction images of coherently illuminated objects in the presence of aberrations," *Opt. Acta* 16, 205-223 (1969).
6. H. Stark (editor), *Image Recovery: Theory and Applications* (Academic, New York, 1987).
7. J. Fienup, "Reconstruction of an object from the modulus of its Fourier transforms," *Opt. Lett.* 3, 27-29 (1978).
8. J. Fienup, "Phase retrieval algorithms," *Appl. Opt.* 21, 2758-2769 (1982). Contains references to many of his previous papers on the Fienup algorithm.
9. R. Barakat and G. Newsam, "Algorithms for reconstruction of partially known, band-limited Fourier transform pairs from noisy data, I. The prototypical linear problem; II. The nonlinear problem of phase retrieval," *J. Integral Eqs.* 9, 49-125 (1985).
10. R. Barakat and G. Newsam, "Necessary conditions for a unique solution to two-dimensional phase retrieval," *J. Math. Phys.* 23, 3190-3193 (1984).

11. G. Newsam and R. Barakat, *Phase Retrieval in Two Dimensions*, Research Report CMA-R15-86, Australian National University Centre for Mathematical Analysis (Canberra, Australia, 1986).
12. N.E. Hurt, *Phase Retrieval and Zero Crossings* (Kluwer, Boston, 1989).
13. R. Barakat and B. Sandler, "Determination of the wavefront aberration function from measured values of the point spread function: A two-dimensional phase retrieval problem," *J. Opt. Soc. Am.*, Series A (in press).
14. A.N. Tichonov and V. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977).
15. Y. A. Morozov, *Methods of Solving Incorrectly Posed Problems* (Springer Verlag, New York, 1984), Chapter 2.
16. J. Dennis and R. Schnable, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1983).
17. R. Fletcher, *Practical Methods of Optimization*, 2nd ed. (Wiley, New York, 1987).
18. R. Barakat and G. Newsam, "Numerically stable iterative method for the inversion of wavefront aberrations from measured point-spread function data," *J. Opt. Soc. Am.* **70**, 1255-1263 (1980).
19. J. J. More, "Recent developments in algorithms and software for trust region methods," in: *Mathematical Programming: The State of the Art*, eds. A. Bachem, M. Grotschel, and B. Korte (Springer-Verlag, Berlin, 1983), pp. 258-287.

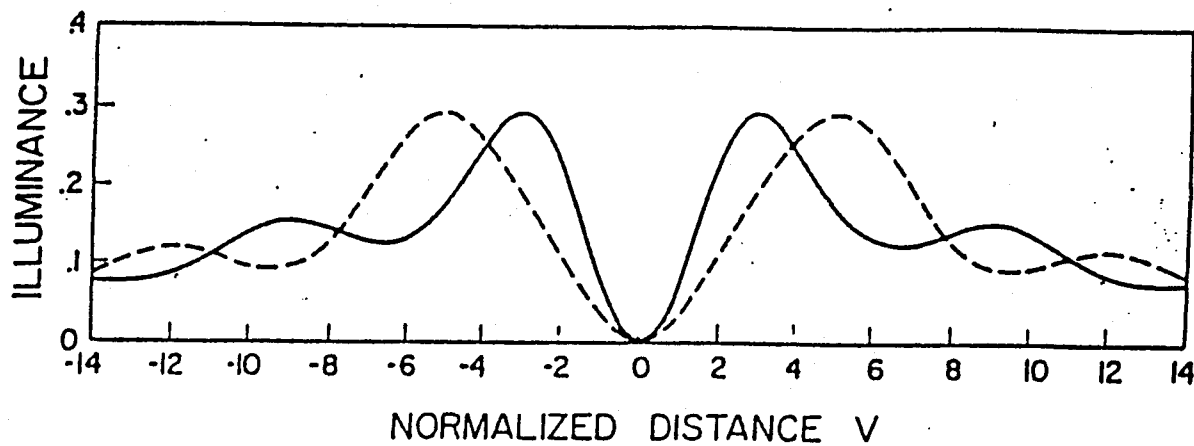


Fig. 1 Distribution of illuminance in the diffraction image of a coherently illuminated opaque edge: $o(v) = 0$, $v < 0$ and $o(v) = 1$ for $v > 0$ viewed through an annular aperture of obscuration radius $\epsilon = 0.05$. The solid line is for the in-focus situation, the dotted line is for a half-wave of defocus. Taken from Reference 5.

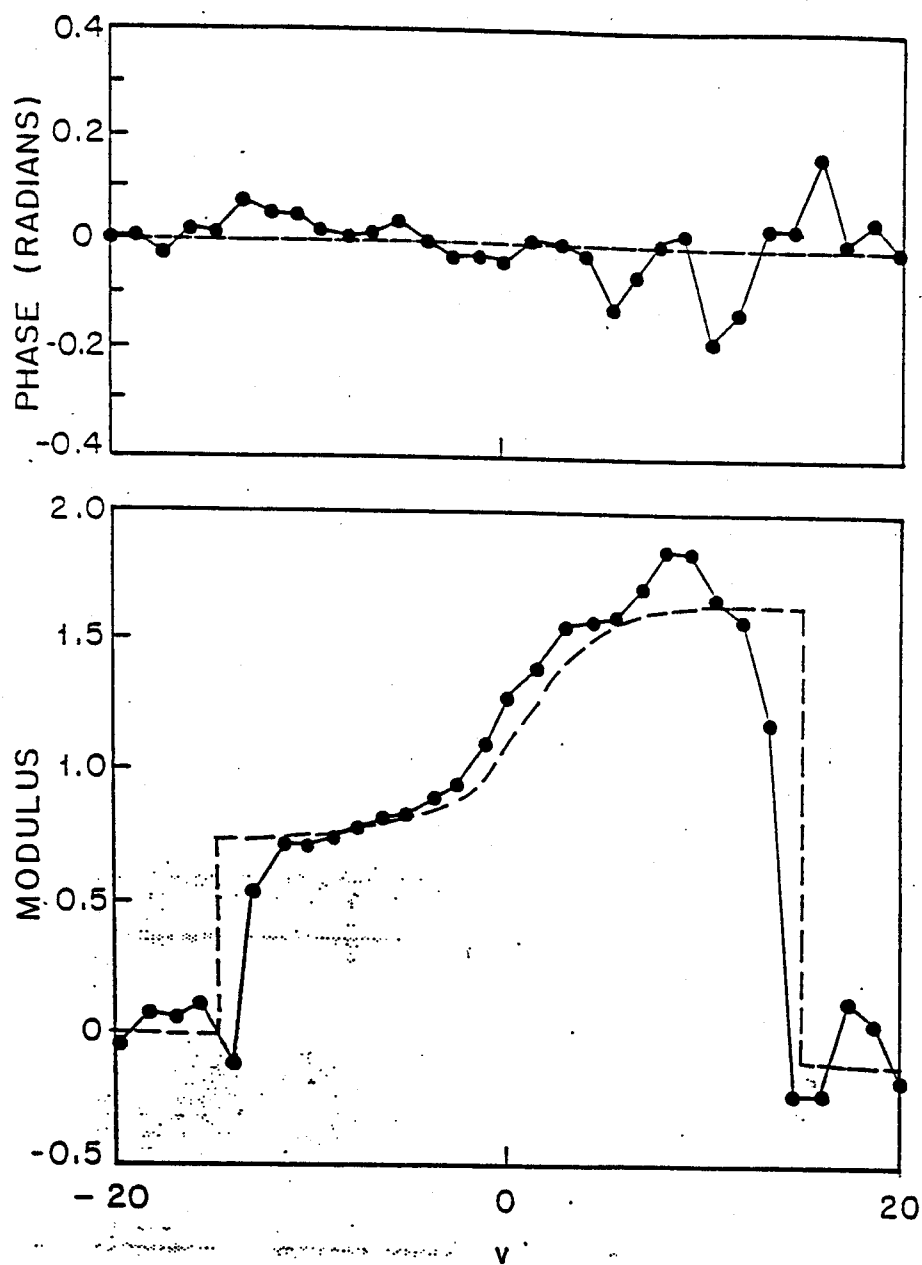


Fig. 2 Sample realization of reconstruction of modulus and phase of coherently illuminated object (dashed lines) in the presence of 2% measurement "noise".

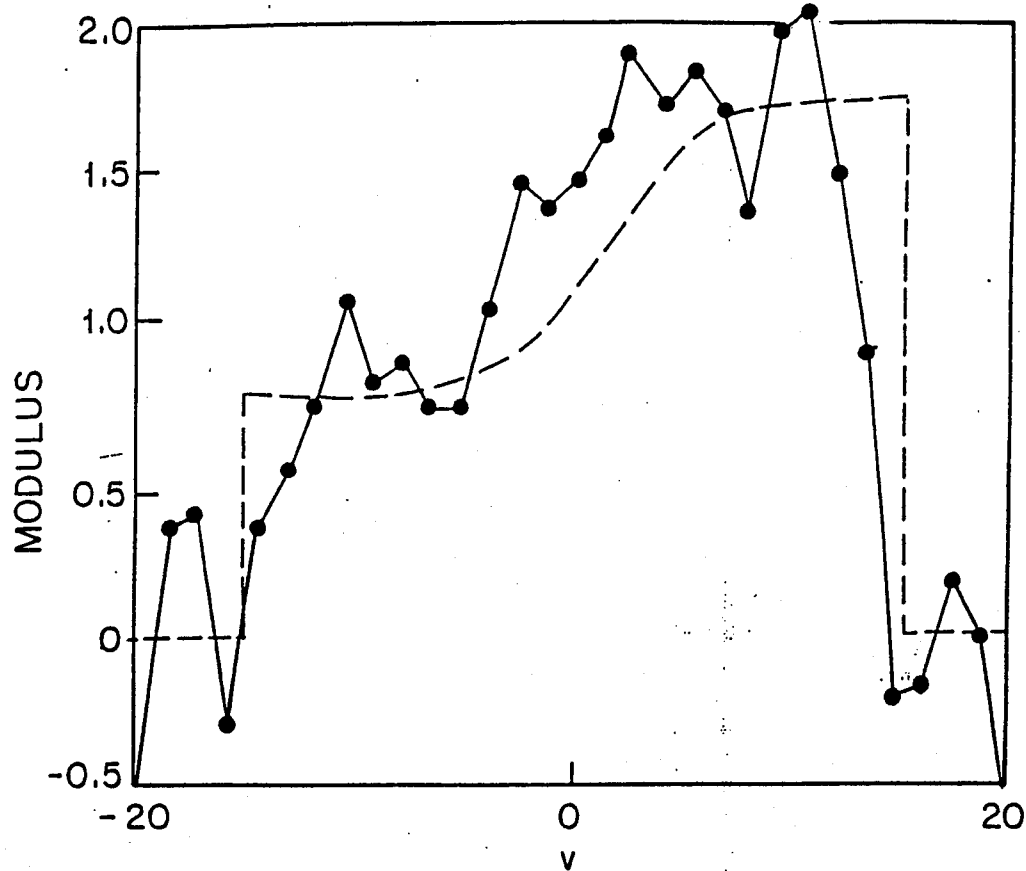


Fig. 3 Sample realization of reconstruction of modulus of coherently illuminated object (dashed line) in the presence of 4% measurement "noise".

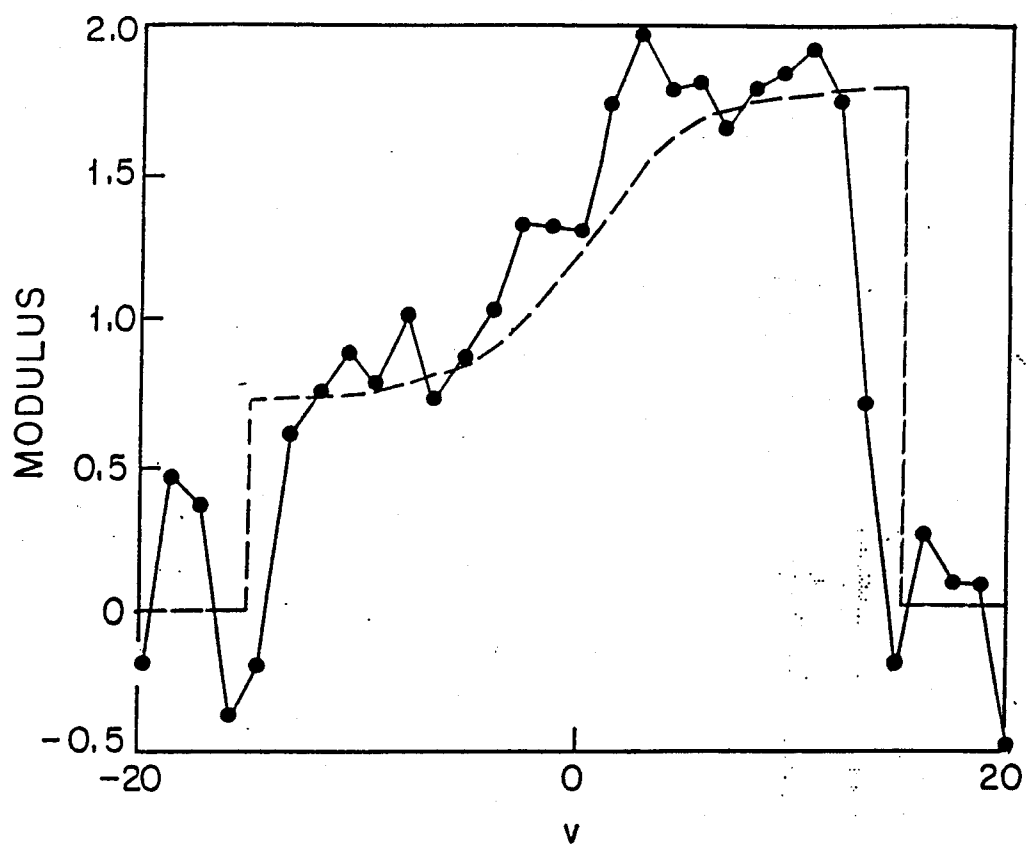


Fig. 4 Sample realization of reconstruction of modulus of coherently illuminated object (dashed line) in the presence of 4% measurement "noise".

REPORT OF INVENTIONS

There were no inventions developed under the aegis of this contract.